A Thesis Title

by

Allison Schneider

Submitted to the Dept. of Earth, Atmospheric and Planetary Sciences
in partial fulfillment of the requirements for the degree of

Bachelor of Science in Earth, Atmospheric and Planetary Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

Author .....
Dept. of Earth, Atmospheric and Planetary Sciences
August 5, 2017

Certified by .....
Glenn R. Flierl
Professor of Oceanography
Thesis Supervisor

Accepted by .....
Richard P. Binzel
Chairman, Committee on Undergraduate Program

A Thesis Title

by

Allison Schneider

Submitted to the Dept. of Earth, Atmospheric and Planetary Sciences
on August 5, 2017, in partial fulfillment of the
requirements for the degree of
Bachelor of Science in Earth, Atmospheric and Planetary Sciences

Abstract

In this thesis, I designed and implemented a compiler which performs optimizations that reduce the number of low-level floating point operations necessary for a specific task; this involves the optimization of chains of floating point operations as well as the implementation of a "fixed" point data type that allows some floating point operations to simulated with integer arithmetic. The source language of the compiler is a subset of C, and the destination language is assembly language for a micro-floating point CPU. An instruction-level simulator of the CPU was written to allow testing of the code. A series of test pieces of codes was compiled, both with and without optimization, to determine how effective these optimizations were.

Thesis Supervisor: Glenn R. Flierl
Title: Professor of Oceanography

Acknowledgments

This is the acknowledgements section. You should replace this with your own acknowledgements.

Contents

List of Figures

List of Tables

1 Introduction

1.1 The Aerocene project

1.2 Acceleration from geopotential height

In a dynamic isobaric trajectory model, the acceleration of a parcel of air at a given pressure level can be determined from its velocity, latitude, and geopotential height. The geopotential height Z_g of a surface of pressure p above mean sea level is

$$Z_g(p) = R \int_p^{p_s} \frac{T}{g} \frac{dp}{p}, \tag{1.1}$$

where R is the gas constant, T is temperature, p_s is surface pressure, g is acceleration due to gravity at mean sea level, and $Z_g(p_s)$ is set to 0. In the troposphere, the difference between geopotential height and actual height is negligible [Marshall and Plumb, 2008].

The geopotential Φ is the potential energy of the Earth's gravitational field at a height h:

$$\Phi(h) = gZ_g(h) \tag{1.2}$$

The full equation for acceleration of an air parcel in the atmosphere is

$$\frac{D\vec{u}}{Dt} + \frac{1}{\rho}\nabla p + \nabla\Phi + f\hat{z} \times \vec{u} = \mathcal{F} \tag{1.3}$$

where $\vec{u}$ is the wind velocity vector, ρ is the density of air, f is the Coriolis parameter, and $\mathcal{F}$ is the force of friction [Marshall and Plumb, 2008]. Due to the isobaric assumption, the gradient of pressure is zero, and the force of friction is assumed to be negligible at the studied pressure level. Using these assumptions and the relation in 1.2, equation 1.3 becomes

$$\frac{D\vec{u}}{Dt} + \nabla\Phi + f\hat{z} \times \vec{u} = 0 \tag{1.4}$$

$$-g\nabla Z_g - f\hat{z} \times \vec{u} = \frac{D\vec{u}}{Dt} \tag{1.5}$$

of which the horizontal components are

$$\frac{Du}{Dt} = fv - g\frac{\partial Z_g}{\partial x} \tag{1.6}$$

$$\frac{Dv}{Dt} = -fu - g\frac{\partial Z_g}{\partial y}. \tag{1.7}$$

2 Methods

Two trajectory calculation routines, a kinematic and a dynamic model, were written using Python's NumPy scientific computing package. Both routines numerically predict the trajectories of an ensemble of parcels by determining the velocities of parcels over time. The kinematic routine finds these velocities by interpolating between grids of wind speed data. The dynamic routine calculates velocities using advection equations relating the parcel acceleration and the geopotential height of a given pressure level.

2.1 Ten day dataset

Data from the Global Forecast System (GFS), a weather forecast model produced by the National Centers for Environmental Protection (NCEP), was used for both models. The dataset chosen was a ten day forecast, starting at 12:00:00 on February 21st, 2017, with predictions at intervals of three hours. Each file in the dataset contains atmospheric predictions for the beginning of a three-hour interval. The values of atmospheric variables are predicted at each point on a latitude-longitude grid spanning the globe, with a spacing between gridpoints of 0.25 degree. Values are predicted for East-West and North-South wind speed components u and v, as well as the geopotential height Z_g of the 250 hectopascal (hPa) pressure level.

2.2 Linear interpolation

Both the kinematic and dynamic models require an interpolation scheme to produce values for atmospheric variables at positions between the gridded values provided by GFS. Linear interpolation is the standard choice for trajectory models [Bowman et al., 2013]. For both models, linear interpolation was used in three dimensions (latitude, longitude, and time). In the kinematic model, u and v components of wind speed were interpolated to the current positions of the air parcels, while in the dynamic model, geopotential height was interpolated.

2.3 Second-order integration scheme

The numerical scheme chosen was a second-order Runge-Kutta method which has a long track record in trajectory modeling [Petterssen, 1940]. The velocity at a given timestep is taken to be the average of the velocity at the initial position and the velocity at the first-guess position after one timestep.

The first guess position $\vec{P}'(t + \Delta t)$ is

$$\vec{P}'(t + \Delta t) = \vec{P}(t) + \vec{V}(\vec{P}, t)\Delta t \tag{2.1}$$

and the final position $\vec{P}(t + \Delta t)$ is

$$\vec{P}(t + \Delta t) = \vec{P}(t) + \frac{1}{2}\left[\vec{V}(\vec{P}, t) + \vec{V}(\vec{P}', t + \Delta t)\right]\Delta t \tag{2.2}$$

where $\vec{P}$ is a position vector with latitude and longitude components, and $\vec{V}$ a velocity vector with u and v wind speeds [Draxler and Hess, 1997]. This integration method is used by HYSPLIT and a number of other trajectory models, including FLEXPART, LAGRANTO, and STILT [Bowman et al., 2013, Stein et al., 2015] For trajectories calculated from interpolated gridded wind velocities, higher order integration schemes do not add precision [Draxler and Hess, 1997].

2.4 Constant timestep

For simplicity, both models employ a constant timestep. To save computation, HYSPLIT uses a dynamic timestep, varying from one minute to one hour, computed to satisfy

$$U_{max}[\text{grid-units min}^{-1}]\Delta t[\text{min}] < 0.75[\text{grid-units}] \tag{2.3}$$

[Draxler and Hess, 1997]. This ensures that the parcel does not blow past any grid squares during a single timestep, which maximizes the accuracy of the calculation. The value of the constant timestep for the experimental models was chosen so that Equation 2.3 is always satisfied. The process of choosing the timestep is detailed in Section 3.3.

2.5 Kinematic equations

At each timestep, after u and v speeds were interpolated and an average value found using the integration scheme, the kinematic model used two equations to solve for a parcel's displacement. Since GFS provides u and v values in meters per second, the equations convert from Cartesian to geographic coordinates. The r value of a parcel is taken to be the radius of the Earth R_E plus the parcel's geopotential height Z_g.

$$r = R_E + Z_g \tag{2.4}$$

$$\frac{d\varphi}{dt} = \frac{v}{r} \tag{2.5}$$

$$\frac{d\lambda}{dt} = \frac{u}{r\cos\varphi} \tag{2.6}$$

The initial conditions for these equations (latitude and longitude) are specified by choosing the initial position of a parcel.

2.6 Dynamic equations

In the dynamic model, velocity at the next timestep was calculated using advection equations which incorporate the current geopotential height gradient and the previous timestep's u and v values. The Coriolis parameter f measures the effect of the Earth's rotation speed Ω at a given latitude φ. Standard acceleration due to gravity is g.

$$f = 2\Omega \sin \varphi \tag{2.7}$$

$$\frac{du}{dt} = fv - \frac{g}{r \cos \varphi} \frac{\partial Z_g}{\partial \lambda} \tag{2.8}$$

$$\frac{dv}{dt} = -fu - \frac{g}{r} \frac{\partial Z_g}{\partial \varphi} \tag{2.9}$$

Equations 2.8 and 2.9 are equations 1.6 and 1.7 transformed from Cartesian to geographic coordinates. After velocity at the next timestep is determined, the dynamic model also uses kinematic equations 2.5 and 2.6 to find the parcel position. This system of four differential equations requires four initial conditions: initial zonal and meridional velocities are needed in addition to initial latitude and longitude. Initial velocity components are set as the geostrophic wind at the initial position,

$$u_g = \frac{-g}{f} \frac{\partial Z_g}{\partial \varphi} \frac{1}{r} \tag{2.10}$$

$$v_g = \frac{g}{f} \frac{\partial Z_g}{\partial \lambda} \frac{1}{r \cos \varphi}. \tag{2.11}$$

2.7 Root-mean-square error measures precision

For an ensemble of parcels, variance among trajectories over time was measured by calculating the mean trajectory: the path of an imaginary parcel whose position at each timestep is the average of the parcels' positions. At each timestep, the root-mean-square error (RMSE) is the square root of the average squared value of each

particle's distance from the mean trajectory.

The mean trajectory was determined by finding the centroids of parcel positions at each timestep after converting trajectory latitudes and longitudes to Cartesian coordinates. NumPy's `arctan2(y,x)` is a two-argument arctangent function with a range of $(-\pi, \pi]$, N is the number of parcels, and n is the parcel index.

$$x_n = \cos\varphi_n \cos\lambda_n \qquad \bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{2.12}$$

$$y_n = \cos\varphi_n \sin\lambda_n \qquad \bar{y} = \frac{1}{N}\sum_{n=1}^{N} y_n \tag{2.13}$$

$$z_n = \sin\varphi_n \qquad \bar{z} = \frac{1}{N}\sum_{n=1}^{N} z_n \tag{2.14}$$

$$\bar{\lambda} = \arctan2\left(\bar{y}, \bar{x}\right) \tag{2.15}$$

$$\bar{\varphi} = \arctan2\left(\bar{z}, \sqrt{\bar{x}^2 + \bar{y}^2}\right) \tag{2.16}$$

The distance d between each parcel (with position φ_n, λ_n) and the mean parcel was calculated with the haversine formula

$$\Delta\varphi_n = |\varphi_n - \bar{\varphi}| \tag{2.17}$$

$$\Delta\lambda_n = |\lambda_n - \bar{\lambda}| \tag{2.18}$$

$$a_n = \sin^2\left(\frac{\Delta\varphi_n}{2}\right) + \cos\varphi_n \cos\bar{\varphi} \sin^2\left(\frac{\Delta\lambda_n}{2}\right) \tag{2.19}$$

$$c_n = 2 \cdot \arctan2\left(\sqrt{a_n}, \sqrt{1 - a_n}\right) \tag{2.20}$$

$$d_n = R_E \cdot c_n \tag{2.21}$$

which is accurate and well-conditioned for small angles [Sinnott, 1984]. The RMSE at each timestep is calculated using the distance between each parcel and the mean parcel.

$$\mathrm{RMSE}\,(t) = \sqrt{\sum_{n=1}^{N} \frac{d_n(t)^2}{N}} \tag{2.22}$$

2.8 Absolute and relative horizontal transport deviation measure accuracy

Because the RMSE uses the mean trajectory as a reference value, it measures variance among trajectories of nearby parcels within a model. To compare trajectories between models, however, previous work has defined the absolute and relative horizontal transport deviation measures (AHTD and RHTD) [Stohl, 1998]. In this study, the AHTD is defined as

$$\mathrm{AHTD}\,(t) = \frac{1}{N}\sqrt{\sum_{n=1}^{N} d_n(t)^2} \tag{2.23}$$

This is similar to the RMSE, but the factor of $\frac{1}{N}$ has been moved outside the square root to agree with the standard definition [Kuo et al., 1985, Rolph and Draxler, 1990]. The values for d_n are calculated as in the RMSE, with the mean latitude $\bar{\varphi}$ and longitude $\bar{\lambda}$ in Equations 2.17 and 2.18 replaced by the latitude and longitude of a reference trajectory. The RHTD is the AHTD normalized by a length L_H.

$$\mathrm{RHTD}\,(t) = \frac{\mathrm{AHTD}\,(t)}{L_H} \tag{2.24}$$

In this study, L_H is defined as in Rolph and Draxler as the mean absolute horizontal travel distance of the reference trajectories,

$$L_H = \frac{1}{N} \sum_{n=1}^{N} \sqrt{\sum_{t=\Delta t}^{T} d_n(t)^2} \tag{2.25}$$

although other definitions of L_H have been used [1990]. Here Δt is the timestep and T is the total trajectory time. Again d_n is calculated as in the RMSE, but with Equations 2.17 and 2.18 replaced by the difference between positions of a single parcel at two different timesteps:

$$\Delta\varphi_n = |\varphi_n(t) - \varphi_n(t - \Delta t)| \tag{2.26}$$

$$\Delta\lambda_n = |\lambda_n(t) - \lambda_n(t - \Delta t)|. \tag{2.27}$$

Although the "true" trajectories of parcels from the ten-day study period are unobservable, reference trajectories which approximate them can be calculated. Reference trajectories were obtained using NOAA's Real-time Environmental Applications and Display sYstem (READY), a tool which uses the HYSPLIT model to generate parcel trajectories [Rolph et al., 2017]. Analysis data from the Global Data Assimilation System (GDAS), with a grid spacing of 0.5 degree and time resolution of three hours, was selected as the source of wind field information [NOAA, 2004]. Trajectories were calculated using HYSPLIT's vertical transport modeling capability, in contrast to the isobaric assumption of this study's kinematic and dynamic models. Fully three-dimensional trajectories and analysis data are more accurate than isobaric trajectories and forecast data, so the reference trajectories will be closer to the "true" parcel trajectories than those calculated by the kinematic and dynamic models. [Stohl, 1998].

3 Tuning the models

3.1 Inertial circles in the dynamic model

3.2 Adding friction to the dynamic model

[[[Inertial circles have been encountered before in dynamic models and there are different ways of dealing with them.]]] [Stohl and Seibert, 1998] One approach, used by Stohl and Seibert, is to take a weighted average of the wind speeds from the dynamic model and interpolated speeds from a kinematic model at each timestep. Another approach is to change the frictionless assumption of the dynamical model. [[[Why does the force of friction damp oscillations?]]] Using the definition of geostrophic wind in Equations 2.10 and 2.11, Equations 2.8 and 2.9 with an added friction term become

$$\frac{du}{dt} = f(v - v_g) - r_f(u - u_g) \tag{3.1}$$

$$\frac{dv}{dt} = -f(u - u_g) - r_f(v - v_g). \tag{3.2}$$

The friction parameter r_f is analogous to the Coriolis parameter: it has units of second^{-1} and is a measure of the force of friction.

Figure 3-1: *Left*: Twenty five trajectories calculated with the frictionless dynamic model. Parcels were launched in an evenly-spaced 5x5 grid with its lower-left corner at 41, -72 and upper-right corner at 42, -71. The large spirals are implausible for jet stream flow and reflect a problem with the model. *Right*: The same parcels with trajectories calculated by the dynamic model with friction. A few parcels still exhibit unrealistic spirals, but in general, trajectories are more plausible.

3.3 Choosing a timestep

[[[test]]]

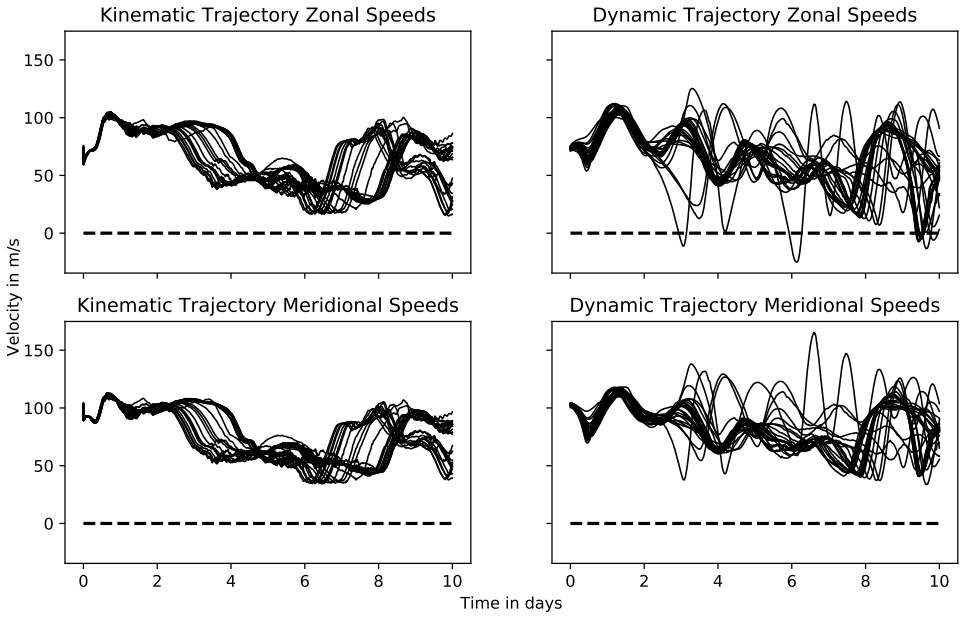

Figure 3-2: For a timestep of 3 minutes, [[[blah blah]]]

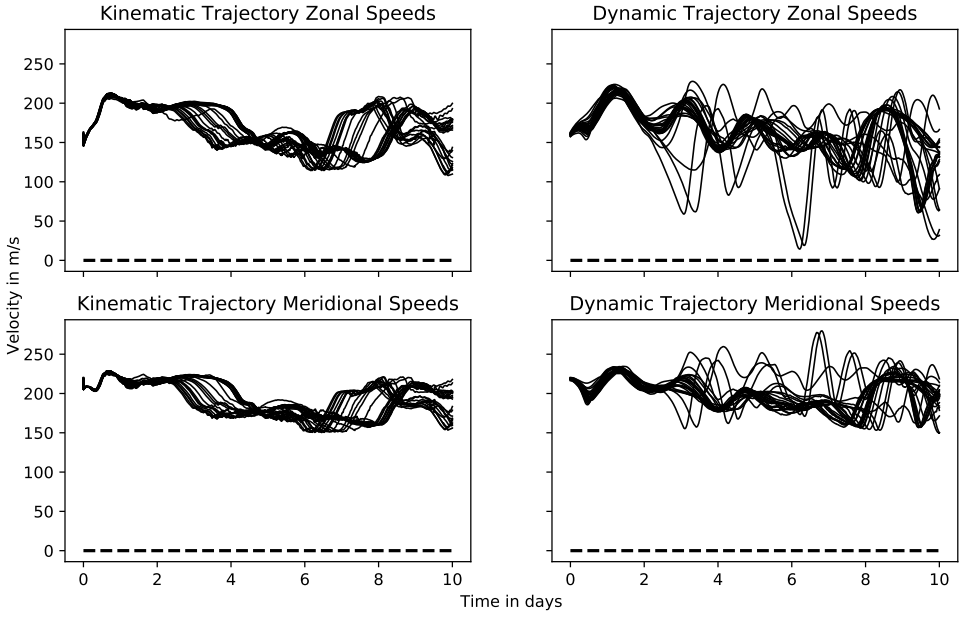

Figure 3-3: For a timestep of 1 minute and 30 seconds, [[[blah blah]]]

4 Results

4.1 Measures of transport error

Table 4.1: Measures of transport error after ten days.

			Launch Site		
Statistic	Model	Reference	Boston	Place 2	Place 3
RMSE	kinematic	mean trajectory	2,630 km		
	dynamic	mean trajectory	2,550 km		
AHTD	kinematic	HYSPLIT 3D			
	dynamic	HYSPLIT 3D			
	kinematic	HYSPLIT isobaric			
	dynamic	HYSPLIT isobaric			
RHTD	kinematic	HYSPLIT 3D			
	dynamic	HYSPLIT 3D			
	kinematic	HYSPLIT isobaric			
	dynamic	HYSPLIT isobaric			
	HYSPLIT isobaric	HYSPLIT 3D			

Bibliography

[Bowman et al., 2013] Bowman, K., Lin, J. C., Stohl, A., Draxler, R., and Konopka, P. (2013). Input data requirements for Lagrangian trajectory models. *Bulletin of the American Meteorological Society*, 94(7):1051–1058.

[Draxler and Hess, 1997] Draxler, R. R. and Hess, G. D. (1997). Description of the HYSPLIT_4 modeling system. Technical report, NOAA Air Resources Laboratory, Silver Spring, MD.

[Kuo et al., 1985] Kuo, Y., Skumanich, M., Haagenson, P. L., and Chang, J. S. (1985). The Accuracy of Trajectory Models as Revealed by the Observing System Simulation Experiments. *Monthly Weather Review*, 113:1852–1867.

[Marshall and Plumb, 2008] Marshall, J. and Plumb, R. A. (2008). *Atmosphere, Ocean, and Climate Dynamics: An Introductory Text*, volume 93 of *International Geophysics Series*. Elsevier Academic Press.

[NOAA, 2004] NOAA (2004). Global Data Assimilation System (GDAS1) Archive Information. Technical report, NOAA-Air Resources Laboratory, Silver Spring, MD.

[Petterssen, 1940] Petterssen, S. (1940). *Weather Analysis and Forecasting*. McGraw-Hill Book Company, New York.

[Rolph and Draxler, 1990] Rolph, G. D. and Draxler, R. R. (1990). Sensitivity of Three-Dimensional Trajectories to the Spatial and Temporal Densities of the Wind Field. *Journal of Applied Meteorology*, 29:1043–1054.

[Rolph et al., 2017] Rolph, G. D., Stein, A. F., and Stunder, B. J. B. (2017). Real-time Environmental Applications and Display sYstem: READY. *Environmental Modelling & Software*, 95:210–228.

[Sinnott, 1984] Sinnott, R. W. (1984). Virtues of the Haversine. *Sky & Telescope*, 68(2):159.

[Stein et al., 2015] Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F. (2015). NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bulletin of the American Meteorological Society*, 96:2059–2077.

[Stohl, 1998] Stohl, A. (1998). Computation, accuracy and applications of trajectories—a review and bibliography. *Atmospheric Environment*, 32(6):947–966.

[Stohl and Seibert, 1998] Stohl, A. and Seibert, P. (1998). Accuracy of trajectories as determined from the conservation of meteorological tracers. *Quarterly Journal of the Royal Meteorological Society*, 124:1465–1484.